

Fordítóprogramok

4. előadás

Aszalós László

2015. október 5.

LL elemzés

Nyelvtan

- $S \rightarrow aS \mid bAc$
- $A \rightarrow bAc \mid d$

Táblázat (M)

	a	b	c	d
S	$(aS, 1)$	$(bAc, 2)$		
A		$(bAc, 3)$		$(d, 4)$

LL elemzés

- Elemzés állapotát egy hármas jelöli: ahol az első elem a még nem elemzett szöveg, a második a mondatforma még nem elemzett része (verem), a harmadik az helyettesítési szabályok sorozata.
- Kezdőállapot $(x$, S, $\lambda)$, ahol $\$$ jelöli a verem, illetve az input végét.$$
- Az elemzés sikeres, ha az utolsó állapot $($, $\$, w)$.$
- Az elemzést segíti egy M elemzőtáblázat $(N \times T)$.
 $M(X, b) = (b\beta, i)$, ha $x \rightarrow b\beta$ az i -dik helyettesítési szabály.
- Állapotátmenetek:
 - ▶ $(ay$, $a\alpha$, $v) \mapsto (y$, α, $v)$ (pop) és$$$$
 - ▶ $(ax$, $Y\alpha$, $v) \mapsto (ax$, $a\beta$, $vi)$, ha $M(Y, a) = (a\beta, i)$.$$$$

Egyszerű LL nyelvtan

A G nyelvtant *egyszerű $LL(1)$ nyelvtannak* nevezzük, ha

- λ -mentes,
- minden helyettesítési szabály jobb oldala terminálissal kezdődik
- alternatívák esetén a jobb oldalak kezdő terminálisai páronként különböznek, azaz

$A \rightarrow a_1\alpha_1 \mid \dots \mid a_k\alpha_k$, ahol $a_i \neq a_j$, ha $i \neq j$

λ -mentes LL(1) nyelvtan

Egy α szimbólumsorozat legbaloldalibb levezetéseiből származó legbaloldalibb terminális szimbólumait jelöljük $FIRST(\alpha)$ -val:

$$FIRST(\alpha) = \{a \mid \alpha \xRightarrow{*} a\beta\}.$$

Definíció

A G λ -mentes nyelvtant λ -mentes LL(1) nyelvtannak nevezzük, ha minden A nemterminális szimbólumra és $k > 1$ -re $A \rightarrow \alpha_1 \mid \dots \mid \alpha_k$ esetén $FIRST(\alpha_i) \cap FIRST(\alpha_j) = \emptyset$, ha $i \neq j$.

λ -mentes LL(1) nyelvtan

Tétel

Ha egy G nyelvtan egyszerű LL(1) nyelvtan, akkor a G egy λ -mentes LL(1) nyelvtan.

Tétel

Ha a G nyelvtan λ -mentes LL(1) nyelvtan, akkor megadható egy vele ekvivalens G' egyszerű LL(1) nyelvtan.

Megjegyzés

Az M elemző táblázat elkészítésekor az az eltérés, hogy $M(X, b) = (\beta, i)$, ha $X \rightarrow \beta$ az i -dik helyettesítési szabály, és $b \in FIRST(\beta)$.

Feladat

Készítse el annak a nyelvtannak az elemző táblázatát, ahol a helyettesítési szabályok:

- $S \rightarrow ABC,$
- $A \rightarrow a|Bbc|Ccd,$
- $B \rightarrow bBb|cCc,$
- $C \rightarrow dDd|Dd$ és
- $D \rightarrow e$

LL(k) nyelvtan

Definíció

Legyen $FIRST_k(\alpha)$ ($k \geq 0$) az α -ból levezethető szimbólumsorozatok k hosszúságú kezdő terminális sorozatainak halmaza, azaz

$$FIRST_k(\alpha) = \{x | \alpha \xRightarrow{*} x\beta \text{ és } |x| = k\} \cup \{x | \alpha \xRightarrow{*} x \text{ és } |x| < k\}.$$

Definíció

A G nyelvtan LL(k) nyelvtan ($k \geq 0$), ha tetszőleges

$$S \xRightarrow{*} wA\beta \Rightarrow w\alpha_1\beta \xRightarrow{*} wx, S \xRightarrow{*} wA\beta \Rightarrow w\alpha_2\beta \xRightarrow{*} wy, \text{ levezetéspárra}$$
$$FIRST_k(x) = FIRST_k(y) \text{ esetén } \alpha_1 = \alpha_2.$$

LL(k) nyelvtan

Tétel

A G nyelvtan akkor és csak akkor LL(k) nyelvtan, ha minden $S \xRightarrow{*} wA\beta$ levezetésre $A \rightarrow \gamma|\delta$ esetén $FIRST_k(\gamma\beta) \cap FIRST_k(\delta\beta) = \emptyset$.

FOLLOW halmaz

Definíció

Legyen $FOLLOW_k(\beta)$ ($k \geq 1$) a β -t tartalmazó mondatformák β utáni szimbólumsorozatainak k hosszúságú terminális prefixeiből álló halmaz, azaz $FOLLOW_k(\beta) = \{x | S \xRightarrow{*} \alpha\beta\gamma \text{ és } x \in FIRST_k(\gamma)\}$, és ha $\lambda \in FOLLOW_k(\beta)$, akkor legyen $FOLLOW_k(\beta) = FOLLOW_k(\beta) \setminus \{\lambda\} \cup \{\$\}$.

Tétel

A G nyelvtan akkor és csak akkor LL(1) nyelvtan, ha minden A nemterminális szimbólumra $A \rightarrow \gamma | \delta$ esetén

$$FIRST_1(\gamma FOLLOW_1(A)) \cap FIRST_1(\delta FOLLOW_1(A)) = \emptyset$$

Erős LL(k) nyelvtan

Definíció

A G nyelvtant erős LL(k) nyelvtannak ($k \geq 0$) nevezzük, ha tetszőleges $S \xRightarrow{*} wA\beta \Rightarrow w\alpha_1\beta \xRightarrow{*} wx$, $S \xRightarrow{*} wA\vartheta \Rightarrow w\alpha_2\vartheta \xRightarrow{*} wy$, levezetéspárra $FIRST_k(x) = FIRST_k(y)$ esetén $\alpha_1 = \alpha_2$.

Tétel

A G nyelvtan akkor és csak akkor erős LL(k) nyelvtan, ha minden $A \rightarrow \gamma|\delta$ esetén $FIRST_k(\gamma FOLLOW_k(A)) \cap FIRST_k(\delta FOLLOW_k(A)) = \emptyset$.

Tétel

A G nyelvtan akkor és csak akkor erős LL(1) nyelvtan, ha LL(1) nyelvtan.

- Nem minden nyelvtan alakítható át LL(1) nyelvtanra.
- Nem eldönthető feladat, hogy az L nyelvhez létezik-e LL(1) nyelvtan.

LL(1) elemző táblázata:

$M(X, b) = (\beta, i)$, ha

- $X \rightarrow \beta$ az i -dik szabály és
 - ▶ $b \in FIRST_1(\beta)$ vagy
 - ▶ $\lambda \in FIRST_1(\beta)$ és $b \in FOLLOW_1(X)$

Konkrét példa

Nyelvtan

- $E \rightarrow E + T \mid T$
- $T \rightarrow T * F \mid F$
- $F \rightarrow (E) \mid i$

Balrekurziómentes változat

- $E \rightarrow TE'$
- $E' \rightarrow +TE' \mid \lambda$
- $T \rightarrow FT'$
- $T' \rightarrow *FT' \mid \lambda$
- $F \rightarrow (E) \mid i$

FIRST és FOLLOW halmazok

- $FIRST(E) = FIRST(T) = FIRST(F) = \{ (, i \}$
- $FIRST(E') = \{ +, \lambda \}$
- $FIRST(T') = \{ *, \lambda \}$
- $FOLLOW(E) = \{ \$,) \}$
- $FOLLOW(E') = \{ \$,) \}$
 - ▶ $E \rightarrow TE' \in P$, így $FOLLOW(E) \subseteq FOLLOW(E')$
- $FOLLOW(T) = \{ \$,), + \}$
 - ▶ $E' \xRightarrow{*} \lambda$, így $FOLLOW(E) \subseteq FOLLOW(T)$
- $FOLLOW(T') = \{ \$,), + \}$
- $FOLLOW(F) = \{ \$,), +, * \}$

Elemzőtáblázat

	+	*	()	i	λ
E	-	-	$E \rightarrow TE'$	-	$E \rightarrow TE'$	-
E'	$E' \rightarrow +TE'$	-	-	$E' \rightarrow \lambda$	-	$E' \rightarrow \lambda$
T	-	-	$T \rightarrow FT'$	-	$T \rightarrow FT'$	-
T'	$T' \rightarrow \lambda$	$T' \rightarrow *FT$	-	$T' \rightarrow \lambda$	-	$T' \rightarrow \lambda$
F	-	-	$F \rightarrow (E)$	-	$F \rightarrow i$	-